HANIF
Journal of Information Systems
https://journal.ilmubersama.com/index.php/hanif

Machine Learning

Application of Data Mining to Determine the Performance of Family Planning Field Officers (PLKB) Using the C4.5 Algorithm

Perdinal Nasution [1*], *Mulkan Azhari* [2]

[1] *Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia*
[2] *Department of Data Science, Faculty of Computer Science and Informatioan Technology, Universitas Muhammadiyah Sumatera Utara, Medan, 20238, North Sumatera, Indonesia*

ARTICLE INFORMATION

Received: Aug 01, 2025
Revised: Aug 11, 2025
Available Online: Oct 16, 2025

KEYWORDS

Data Mining, C4.5 Algorithm, PLKB Performance (Family Planning Field Officers), Performance Evaluation

CORRESPONDENCE (*)

E-mail: bkkperdi@gmail.com

A B S T R A C T

The effectiveness of family planning programs is closely related to the performance of Family Planning Field Officers (PLKB). Conventional performance evaluation methods often rely on manual assessments, which may lead to subjectivity and inconsistency. To overcome this issue, data mining techniques can be applied to analyze performance data systematically and objectively. This study employs the C4.5 decision tree algorithm to classify and evaluate the performance of PLKB. The dataset used in this research includes several indicators, such as service coverage, counseling frequency, reporting accuracy, and community participation. Prior to model construction, data preprocessing was performed to handle missing values and normalize attributes. The model performance was evaluated using accuracy, precision, recall, and F-measure. The findings indicate that the C4.5 algorithm successfully classified PLKB performance into three categories: high, medium, and low. The model achieved an accuracy of [insert % if available], demonstrating its effectiveness in identifying key determinants of officer performance. Moreover, the decision tree generated interpretable rules that highlight the most influential attributes affecting PLKB performance. The application of data mining using the C4.5 algorithm provides an objective and efficient method for evaluating PLKB performance. This approach not only enhances decision-making for supervision and training but also contributes to the improvement of family planning program implementation. Future research is suggested to compare the C4.5 algorithm with other classification methods to achieve higher accuracy and generalizability.

INTRODUCTION

Employee performance assessments conducted by the Family Planning (KB) Extension Office in Timbang Lawan Village, Namorambe District, cannot be done haphazardly. Therefore, employee performance assessment variables are needed to ensure more accurate assessment results. Previously, employee performance assessments at the Family Planning (KB) Extension Office in Timbang Lawan Village, Namorambe District, were conducted manually using only assessment forms. Therefore, it was deemed necessary to analyze and classify employee performance at the Family Planning (KB) Extension Office in Timbang Lawan Village, Namorambe District. This employee performance assessment used the C4.5 Algorithm data mining approach [1,2,3].

Data mining has emerged as a powerful tool for extracting useful patterns and knowledge from large datasets. Among various classification techniques, the C4.5 decision tree algorithm is widely used because of its ability to generate interpretable models and handle diverse attributes effectively. This makes it suitable for assessing PLKB performance, as

https://doi.org/10.56211/hanif.v3i1.52
Attribution-ShareAlike 4.0 International Some rights reserved

it can classify officers into performance categories while also identifying the most influential factors contributing to their achievements or shortcomings [4,5,6].

The objective of this study is to apply the C4.5 algorithm to classify the performance of PLKB based on multiple indicators, such as service coverage, counseling frequency, reporting accuracy, and community engagement. By implementing this approach, the research aims to provide a systematic and objective evaluation model that supports decision-making in family planning program management. Furthermore, the findings are expected to offer insights for designing better supervision, training, and resource allocation strategies to improve the overall effectiveness of family planning initiatives.

The application of data mining technology is an attractive solution for improving the performance evaluation of Family Planning Field Officers (PLKB) in Timbang Lawan Village. Using the C4.5 algorithm, various data related to PLKB activities and achievements can be analyzed. This data includes the number of household visits, participation rates in outreach activities, and the number of families receiving family planning services, among others. This allows for the generation of predictive models capable of identifying the factors most influential on PLKB performance and determining recommendations for improvement.

This research is expected to provide a significant contribution in the development of a more objective and efficient performance evaluation system for PLKB (Family Planning Field Officers). By utilizing Data Mining technology and the C4.5 algorithm, it is expected to create a system that is able to provide a deeper understanding of the factors that influence the performance of PLKB (Family Planning Field Officers), so that related parties can carry out appropriate and effective interventions in increasing the efficiency and effectiveness of the KB (Family Planning) program at the local level, especially in Timbang Lawan Village, Namorambe District.

METHOD

C4.5 Algorithm

At this stage, the data selected according to research requirements is then processed manually in MS Excel using the C4.5 Algorithm formulas. The formulas used in the data processing calculations using the algorithm are as follows:

1. Entropy
   Entropy (S) is the estimated number of bits needed to extract a class (+ or -) from a random data set in the sample space S. Entropy can be defined as the number of bits required to represent a class.
2. Gain
   Gain (S,A) is the information gain from attribute A relative to the output data S. Information gain is obtained from the output data or dependent variable S grouped based on attribute A, denoted by gain (S,A).
3. Split Info
   Split Info is a formula used to separate a number of data attributes that have been generated from the gain value.
4. Gain ratio
   Gain ratio is a formula used to determine where the next calculation is recalculated again. After obtaining the values from these calculations, you can create a decision tree from them. Decision trees can be interpreted as a very powerful way to predict or clarify. Decision trees can divide large data sets into smaller sets of records by applying a series of decision rules.

Database

MySQL databases are based on the relational data model, which is the foundation of this database management system. In the relational data model, data is organized into tables with rows and columns, where each table represents an entity and each column represents an attribute.

The use of primary keys and foreign keys is a key concept in designing relationships between tables that enable efficient data integration. Other database theories, such as normalization, transactions, and query optimization, also play an important role in the use of MySQL in the context of designing, managing, and using relational databases.

MySQL is a server that serves databases. To create and process databases, we can learn specialized programming called SQL queries (commands). The database itself is needed if we want to input user data using HTML forms, which can then be processed using PHP to be saved into a MySQL database [7,8,9].

Website

A website is a collection of web pages accessible via the internet. Websites consist of various elements, including text, images, videos, and various other types of multimedia content. Developing and maintaining a website often involves specific software and tools, as well as server hardware to run it. To develop and test a website locally before publishing it online, many web developers use tools such as XAMPP and Sublime Text.

XAMPP is a tool that provides software packages in a single package. By installing XAMPP, there is no longer a need to manually install and configure the Apache, PHP, and MySQL web servers. XAMPP will install and configure them automatically. XAMPP is one of the instant installation packages for Apache, PHP, and MySQL that can be used to assist the process [10,11,12].

RESULTS AND DISCUSSION

The interface display results are the stage where the system or application is ready to be operated in actual conditions according to the results of the analysis and design carried out, so that it will be known whether the system or application built can produce a goal that is achieved, and this data mining application is equipped with a display that aims to make it easier for its users. The function of this interface is to provide input and display the output of the application. In this application, it has an interface consisting of a login menu, Community, and a process menu.

Main Menu

The Main Menu displays the initial menus on the system, namely the login menu and the main menu. The Main Menu menu is as follows:

1. Login Menu
   The login menu is used to secure the system from unauthorized users before accessing the main menu. The login menu displays as follows:

Figure 1. Login Menu

2. Main Menu
   The Main Menu serves as a hub for community data, processes, and reports. The main menu displays as follows:

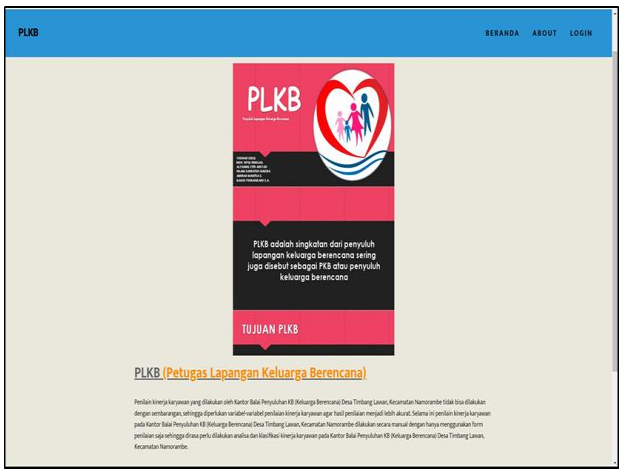

Figure 2. Main Menu

Administrator Page

The administrator page displays the data processing menus for storing data into the database, namely the Community menu and the C4.5 process. The main administrator page menus are as follows:

1. Community Data Menu

   The Community Data menu is for processing community data in Timbang Lawan Village, Namorambe District. The Community Data menu is as follows.

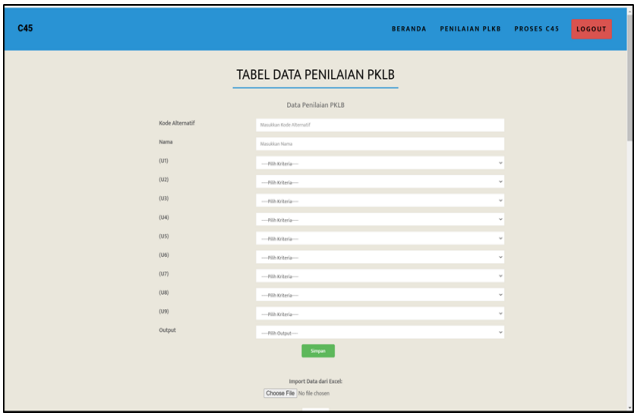

Figure 3. Community Data Menu

2. Process Menu C4.5

   The C4.5 process menu is used to process data to determine the performance of the PKLB in Timbang Lawan Village, Namorambe District, using the Decision Tree method. The C4.5 process menu is as follows.

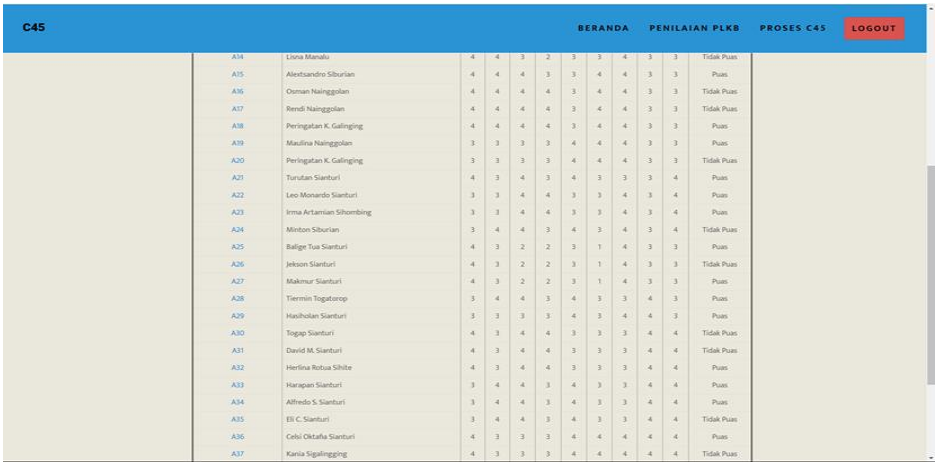

Figure 4. Process Menu C4.5

Testing

In this section, you will be asked to conduct testing using new data samples. In this section, you will be asked to verify the accuracy of the system you designed using previously tested and calibrated tools. The results of the program process in determining PKLB performance are as follows.

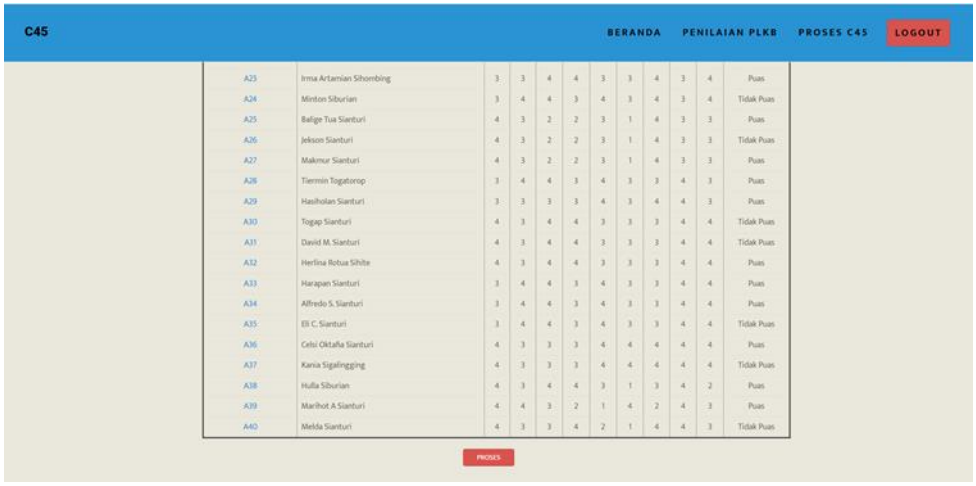

Figure 5. C4.5 Associating Results

Implementation of C4.5

The algorithm process carried out using the C4.5 method is as follows:

1. Gain and Entropy Calculation

   The decision tree is created after calculating the total entropy, using the entropy of each attribute and calculating the gain and determining the highest gain. The total entropy calculation in Table 4.1 can be calculated using the following equation:

   From Table 1, it can be seen that in the total row, the number of Questionnaires (S) is 40, Dissatisfied (TP) is 16, and Satisfied (P) is 24. The total entropy calculation is shown in Table 4.1. can be calculated using the following equation:

   Entropy (Total) = (-24/40 x Log2(24/40)) + (-16/40 x Log2(16/40))

   = 0.97095

   The entropy calculation for each attribute is the same as the total entropy calculation. The gain value for the attribute Gain is calculated using the following equation:

   Gain = 0.97075 ((17/0.97075*0.97742)) + (23/0.97095*0.965664)

   = 0.0003

Table 1. Node 1 Calculation

Factor	Value	Results		TOTAL	ENTROPY	GAIN
		P	TP			
		24	16	40	0.97095	
U1						0.0003
	3	10	7	17	0.977418	
	4	14	9	23	0.965636	
U2						0.0075

 https://doi.org/10.56211/hanif.v3i1.52

	3	11	9	20	0.992774
	4	13	7	20	0.934068
					0.2012
	2	0	5	5	0
U3	3	9	6	15	0.970951
	4	15	5	20	0.811278
					0.0898
	2	1	4	5	0.721928
U4	3	12	4	16	0.811278
	4	11	8	19	0.981941
					0.1530
	1	0	2	2	0
	2	0	1	1	0
U5	3	10	9	19	0.998001
	4	14	4	18	0.764205
					0.278007
	1	0	7	7	0
U6	3	13	4	17	0.787127
	4	11	5	16	0.896038
					0.037495
	2	0	1	1	0
U7	3	11	8	19	0.981941
	4	13	7	20	0.934068
					0.431937
	1	0	6	6	0
	2	0	4	4	0
U8	3	14	3	17	0.672295
	4	10	3	13	0.77935
					0.072242
	1	0	1	1	0
U9	2	0	1	1	0
	3	12	8	20	0.970951
	4	12	6	18	0.918296

Calculating the entropy for each attribute is similar to calculating the gain value. Once the entropy value is obtained, the gain value for each attribute is calculated, as seen in the node 1 calculation. The table above shows that service security has the highest gain. Here is the temporary decision tree:

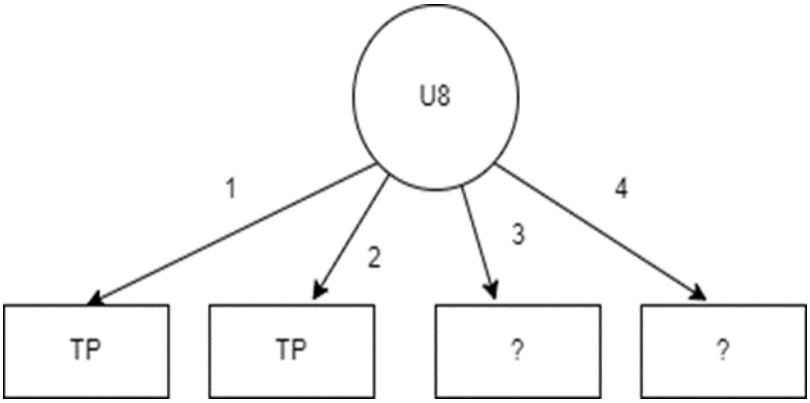

Figure 6. Root Node

The root node above doesn't yet show a dominant decision for each selected statement. Therefore, we must re-examine the entropy and gain values for each Service Security attribute (criterion).

Table 2. Service Security Attributes (Criteria)

Service Elements	Satisfied	Dissatisfied	Total
Service Security (U8)	14	3	17

It can be seen that the Service Security attribute has an entropy value of 0.672295, from this value it still requires further calculations and the results from the node obtained the Fairness of Getting Service attribute has a gain value of 0.0996. From this table it still requires further calculations until the gain and entropy values are the same.

The resulting decision tree provides an explanation, showing that the decision tree has categorized all cases. After training both models (C4.5), you can analyze the prediction results and compare their performance. Be sure to choose the model that provides the most accurate results and aligns with the research objectives, as shown in Figure 7. below.

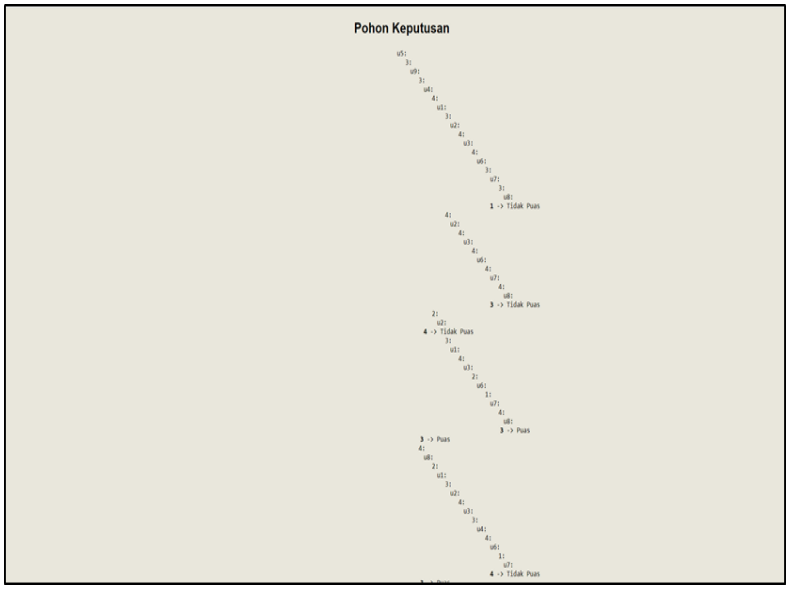

Figure 7. C4.5 Decision Tree Results

 https://doi.org/10.56211/hanif.v3i1.52

CONCLUSION

Based on the analysis of the problems encountered in the case discussed, which focused on determining the level of public satisfaction in determining the performance of the PLKB, the following conclusions can be drawn: Based on the research results, the C4.5 method was applied by implementing the method steps and determining the level of public satisfaction in determining the performance of the PLKB. Based on the research results, the data mining system application was designed by modeling the system and building a web-based system to determine the level of public satisfaction in determining the performance of the PLKB using the C4.5 method. Based on the research results, system testing was conducted by logging into the system and processing variable and process data to determine the level of public satisfaction in determining the performance of the PLKB, by displaying a report on the results of the level of public satisfaction in determining the performance of the PLKB.

REFERENCES

Buku

[1] Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.
[2] Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.
[3] Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.
[4] Janner Simarmata Arsan Kumala Jaya, Syarifah Fitrah Ramadhani, Niel Ananto, Abdul Karim, Betrisandi, Muhammad Ilham Alhari, Cucut Susanto, Suardinata, Indah Purnama Sari, Edson Yahuda Putra. Komputer dan Masyarakat. Medan: Yayasan Kita Menulis, 2024, pp.162.
[5] Mahdianta Pandia, Indah Purnama Sari, Alexander Wirapraja Fergie Joanda Kaunang, Syarifah Fitrah Ramadhani Stenly Richard Pungus, Sudirman, Suardinata Jimmy Herawan Moedjahedy, Elly Warni, Debby Erce Sondakh. Pengantar Bahasa Pemrograman Python. Medan : Yayasan Kita Menulis, 2024, pp.180
[6] Zelvi Gustiana Arif Dwinanto, Indah Purnama Sari, Janner Simarmata Mahdianta Pandia, Supriadi Syam, Semmy Wellem Taju Fitrah Eka Susilawati, Asmah Akhriana, Rolly Junius Lontaan Fergie Joanda Kaunang. Perkembangan Teknologi Informatika. Medan: Yayasan Kita Menulis, 2024, pp.158

Jurnal

[7] Londa, G. O., Witi, F. L., & Bhae, B. Y. (2022). Sistem Informasi Pendataan Penduduk Desa Detusoko Barat Kecamatan Detusoko Kabupaten Ende Berbasis Web. JURNAL JITEK, II(2).
[8] Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve BayesData Preprocessing. JURNAL ILMIAH INFORMATIKA, 75-81.
[9] Amna, S, W., Sudipa, I. I., E. Putra, T. A., Wahidin, A. J., Syukrilla, W. A., . . . Santoso, L. W. (2023). DATA MINING. Sumatera Barat: PT GLOBAL EKSEKUTIF TEKNOLOGI.
[10] Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. Jurnal TEKNO KOMPAK, XVI(2), 29-43.
[11] Astuti, I. P. (2019). Algoritma Apriori Untuk Menemukan Hubungan Antara Jurusan Sekolah Dengan Tingkat Kelulusan Mahasiswa. JURNAL TEKNIK INFORMATIKA, 69-78.
[12] Bachtiar, L., & Mahradianur. (2023). Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai. JURNAL INFORMATIKA, X(1), 28-36.
[13] Budiman, I., Saori, S., Anwar, R. N., Fitriani, & Yuga, M. (2021). Analisis Pengendalian Mutu Di Bidang Industri Makanan (Studi Kasus: UMKM Mochi Kaswari Lampion Kota Sukabumi). Jurnal Inovasi Penelitian, X(1).
[14] Hidayah, A. Z., & Rozi, A. F. (2021). Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus : Universitas Mercu Buana Yogyakarta). JURNAL INFORMATION SYSTEM & ARTIFICIAL INTELLIGENCE, I(2), 117-127.
[15] Juni Arta, I. K., Indrawan, G., & Dantes, G. R. (2019). Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di STMIK Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution. Jurnal Ilmu Komputer Indonesia (JIKI), 11-21.
[16] Sari, I.P., Hariani, P.P., Al-Khowarizmi, A., Ramadhani, F., Sulaiman, O.K., Satria, A, & Manurung, A.A. (2024). CLUSTERING HIV/AIDS DISEASE USING K-MEANS CLUSTERING ALGORITHM. Proceeding International Seminar on Islamic Studies 5 (1), 1668-1676
[17] Mardi, Y. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Data Mining : Klasifikasi Menggunakan Algoritma C4.5, 213-219.

[18] Sari, I.P., Ramadhani, F., Satria, A., & Sulaiman, O.K. Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. 2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 486-491

[19] Manurung, A.A., Nasution, M.D., & Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. 2023 11th International Conference on Cyber and IT Service Management (CITSM), 1-4

[20] Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

[21] Sari, I.P., Al-Khowarizmi, A, Sulaiman, O.K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2), 402-412

[22] Sulaiman, O.K & Batubara, I.H. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 128-135

[23] Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. International Journal of Economic, Technology and Social Sciences (Injects) 2 (1), 331-338

[24] Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering 2 (1), 139-144

[25] Apdilah, D., & Sari, I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 156-163